# The Great Internet TCP Congestion Control Census

Ayush Mishra
ayush@comp.nus.edu.sg
National University of Singapore
Singapore

Xiangpeng Sun
sun.xiangpeng@comp.nus.edu.sg
National University of Singapore
Singapore

Atishya Jain
atishya.jain.cs516@cse.iitd.ac.in
Indian Institute of Technology, Delhi
India

Sameer Pande
sameer.vivek.pande.cs117@cse.iitd.ac.in
Indian Institute of Technology, Delhi
India

Raj Joshi
rajjoshi@comp.nus.edu.sg
National University of Singapore
Singapore

Ben Leong
benleong@comp.nus.edu.sg
National University of Singapore
Singapore

## ABSTRACT

In 2016, Google proposed and deployed a new TCP variant called BBR. BBR represents a major departure from traditional congestion control as it uses estimates of bandwidth and round-trip delays to regulate its sending rate. BBR has since been introduced in the upstream Linux kernel and deployed by Google across its data centers. Since the last major study to identify TCP congestion control variants on the Internet was done before BBR, it is timely to conduct a new census to give us a sense of the current distribution of congestion control variants on the Internet. To this end, we designed and implemented *Gordon*, a tool that allows us to measure the congestion window (cwnd) corresponding to each successive RTT in the TCP connection response of a congestion control algorithm. To compare a measured flow to the known variants, we created a localized bottleneck and introduced a variety of network changes like loss events, changes in bandwidth and delay, while normalizing all measurements by RTT. We built an offline classifier to identify the TCP variant based on the cwnd trace over time.

Our results suggest that CUBIC is currently the dominant TCP variant on the Internet, and is deployed on about 36% of the websites in the Alexa Top 20,000 list. While BBR and its variant BBR G1.1 are currently in second place with a 22% share by website count, their present share of total Internet traffic volume is estimated to be larger than 40%. We also found that Akamai has deployed a unique loss-agnostic rate-based TCP variant on some 6% of the Alexa Top 20,000 websites and there are likely other undocumented variants. Therefore, the traditional assumption that TCP variants "in the wild" will come from a small known set is not likely to be true anymore. Our results suggest that some variant of BBR seems poised to replace CUBIC as the next dominant TCP variant on the Internet.

## CCS CONCEPTS

• **Networks** → **Transport protocols**; Public Internet; • **General and reference** → *Measurement*;


*SIGMETRICS '20 Abstracts, June 8–12, 2020, Boston, MA, USA*

ACM ISBN 978-1-4503-7985-4/20/06.
https://doi.org/10.1145/3393691.3394221

## KEYWORDS

congestion control; measurement study

**ACM Reference Format:**
Ayush Mishra, Xiangpeng Sun, Atishya Jain, Sameer Pande, Raj Joshi, and Ben Leong. 2020. The Great Internet TCP Congestion Control Census. In *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '20 Abstracts), June 8–12, 2020, Boston, MA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3393691.3394221

## 1 INTRODUCTION

Over the past 30 years, TCP congestion control has evolved to adapt to the changing needs of the users and to exploit improvements in the underlying network. Most recently, in 2016, Google proposed and deployed a new TCP variant called BBR [2]. BBR represents a major departure from traditional congestion-window-based congestion control. Instead of using packet loss as a congestion signal, BBR uses estimates of the bandwidth and round-trip delays to regulate its sending rate. BBR has since been introduced in the Linux kernel [1] and deployed by Google across its data centers. As the TCP ecosystem today has changed significantly since the last study done in 2011 [4, 5], it is timely to conduct a new census to understand the latest distribution of TCP variants on the Internet.

Our census aims to (i) understand how the distribution of previously identified variants has changed since the last measurement study in 2011 [4, 5]; (ii) develop a method to identify BBR in existing websites; and (iii) determine the proportion of undocumented TCP variants if any. The final goal of our approach represents a significant departure from previous studies, which assumed a fixed set of known TCP variants and attempted to classify all the measured websites as one of the known variants.

To this end, we designed *Gordon*, a tool that allows us to measure the size of the congestion window (cwnd) corresponding to each successive RTT in the TCP connection response of a congestion control algorithm "in the wild." While rate-based TCP variants do not maintain a congestion window, they typically maintain a maximum allowable number of packets in flight [2], which we can measure as the *effective* congestion window for each RTT. To compare this response to that of known variants, we created a localized bottleneck where we introduced a variety of network changes: loss events, bandwidth change, and increased delay. An offline classifier identified the TCP variant based on this cwnd trace over time.

Our relatively modest design goals introduce a set of unique challenges. Previous studies that also try to measure the cwnd profiles of senders on the Internet [4] utilized delayed ACKs to 'space out' individual cwnds in a packet trace. However, this strategy would not work in measuring the cwnd of rate-based variants that pace packets, like BBR [2]. Also, most of our target web servers host small web pages, which makes it difficult to measure the entire cwnd evolution of its congestion controller. Finally, our tool also needs to eliminate the error from short-term variations in the network conditions on the Internet.

Gordon measures the cwnd of a sender by withholding ACKs instead of delaying them. Since the cwnd of a sender is the maximum number of unacknowledged packets allowed, Gordon first starts a connection, drops all the packets, and counts the number of packets received before it sees a re-transmission. We refer to this as the first cwnd $C_1$. Gordon then connects again and in the next connection, sends data packets in response to the first ACKs back $C_1$ packets, before withholding ACKs again to measure $C_2$. With this strategy, Gordon can measure the entire cwnd response of a sender over multiple connections. Gordon also uses a localized bottleneck to isolate bandwidth changes on the Internet and impose its own bandwidth "profile". This leaves our measurement strategy susceptible only to random packet losses, which Gordon overcomes by making multiple measurements to eliminate any measurement noise. To address the issue of small webpages, we crawled target websites for the largest webpage available on the site and used small MTU sizes to increase the number of packets received from a single download.

We used Gordon to measure the 20,000 most popular websites according to the Alexa rankings [3]. The following are our key findings:

(1) Our results suggest that, as expected, CUBIC is currently the dominant TCP variant on the Internet and is deployed at 36% of all the classified websites, which is an increase from what was reported in the last study in 2011.
(2) The rate of BBR adoption over the past 3 years since its release has been phenomenal. BBR (together with its Google variant BBR G1.1) is currently the second most popular TCP variant deployed at 22% of the classified websites.
(3) While BBR has a share of only 22% by website count, we estimate that its present share of total Internet traffic volume already exceeds 40%. This proportion will almost certainly exceed 50% if Netflix and Akamai also decide to adopt BBR.
(4) The assumption that TCP variants "in the wild" will come from a known set is unlikely to be true anymore. In particular, we found that Akamai has deployed a unique loss-agnostic rate-based TCP variant on about 6% of the Alexa Top 20,000 websites.

Our results are summarized in Table 1. Short-flows are websites that could not sustain connections long enough to be classified due to small page sizes. We also found a small number of websites that were unresponsive to our measurement strategy and terminated connections early. These are referred to as 'Unresponsive' hosts in Table 1.

**Table 1: Distribution of TCP variants used by the Alexa Top 20,000 websites [3] on the Internet.**

| | Variant | Websites | Proportion |
|---|---|---|---|
| Loss-based AIMD | New Reno [r] | 160 | 0.80% |
| Loss-based MIMD | HSTCP [r] | | |
| | CUBIC | 6,139 | 30.70% |
| | BIC | 181 | 0.90% |
| | Scalable | 39 | 0.20% |
| Delay-based AIMD | Vegas [v] | 564 | 2.82% |
| | Westwood | 0 | 0% |
| Delay-based MIMD | CTCP [c] | 1,148 | 5.74% |
| | Illinois [c] | | |
| | Veno [v] | | |
| | YeAH | 1,162 | 5.81% |
| | HTCP | 560 | 2.80% |
| Rate-based | BBR | 3.550 | 17.75% |
| | BBR G1.1 | 167 | 0.84% |
| | AkamaiCC | 1,103 | 5.51% |
| Unknown | | 2,432 | 12.16% |
| Short-flows | | 1,493 | 7.47% |
| Unresponsive | | 1,302 | 6.51% |
| Total hosts | | 20,000 | 100% |

[r] New Reno and HSTCP are classified together.
[v] Vegas and Veno are classified together.
[c] CTCP and Illinois are classified together.

## 2 ACKNOWLEDGMENTS

We thank our shepherd, Michael Sirivianos and the anonymous SIGMETRICS reviewers for their valuable feedback and comments. We thank Neal Cardwell, Yuchung Chen, Nimantha Baranasuriya, Venkat Padmanabhan, Daniel Havey, Praveen Balasubramaniam, Igor Lubashev, and Mike Afergan, for their suggestions and help in corroborating some of the findings presented in this paper.

## 3 RESOURCES

Our measurement tool, along with the cwnd traces for the Alexa Top 20,000 websites is available on GitHub (https://github.com/NUS-SNL/Gordon).

## REFERENCES

[1] Neal Cardwell. 2017. tcp_bbr: add BBR congestion control. (2017). https://bit.ly/2VAJcDD
[2] Neal Cardwell, Yuchung Cheng, C. Stephen Gunn, Soheil Hassas Yeganeh, and Van Jacobson. 2017. BBR: Congestion-based Congestion Control. *CACM* 60, 2 (2017), 58–66.
[3] Alexa Internet Inc. 2018. The Top 500 websites on the Internet. (2018). https://www.alexa.com/topsites
[4] Peng Yang, Juan Shao, Wen Luo, Lisong Xu, Jitendra Deogun, and Ying Lu. 2011. TCP Congestion Avoidance Algorithm Identification. *IEEE/ACM Transactions on Networking* 22, 4 (2011), 1311–1324.
[5] Peng Yang and Lisong Xu. 2011. A survey of deployment information of delay-based TCP congestion avoidance algorithm for transmitting multimedia data. In *Proceedings of GLOBECOM Workshops*.